Circuit and CAD Solutions for optimal SRAM Design in Nanoscale CMOS

Satyanand Nalam
Department of Electrical and Computer Engineering
University of Virginia

A Dissertation Proposal Presented in Partial Fulfillment of the
Requirement for the
Doctor of Philosophy Degree
in Computer Engineering
April, 2010

Abstract

Conventional 6T SRAM design involves balancing trade-offs among several critical metrics - yield, power, performance and area. Reducing noise margins and worsening variation makes scaling the 6T bitcell to newer technologies and lower supply voltage difficult, especially due to the need to balance the trade-offs between various metrics. In particular, lowering the SRAM voltage for low-power operation, while maintaining functionality, is a challenge. Several solutions have been proposed to ensure continued scaling of SRAM in the sub-45nm regime. These include circuit solutions such as alternative bitcells and read/write assist techniques, technology solutions such as new materials and devices, and architectural solutions such as ECC and redundancy. Exploring this vast design space to zero in on an optimal design that is the most suited to a designer's requirements thus becomes challenging.

To deal with the problem of scaling SRAM to sub-45nm technologies, we first propose two alternative bitcell options, a 5T and an asymmetric 6T, that provide more flexibility in exercising these trade-offs. These bitcells exploit asymmetric sizing and single-ended operation to enable more flexible trade-offs between the critical metrics. Next, to enable early design space exploration, we propose a Virtual Prototyping tool (ViPro). The designer uses ViPro to produce virtual prototypes of the SRAM and evaluate the trade-offs between various metrics without requiring a complete implementation of the component circuits. This helps him/her iteratively reach a completed design that is best suited to his/her requirements in terms of the critical metrics. Finally, we propose to investigate dynamic measures of bitcell stability and gain insight into their correlation with conventional static noise margin metrics that are typically used to predict the SRAM minimum operating voltage (V_{MIN}). Since static metrics are either too optimistic or pessimistic, particularly at high frequencies of operation typical of nanoscale SRAMs, knowledge about the correlation or lack thereof between static and dynamic metrics can help better predict the V_{MIN}.

1. Introduction

1.1 Motivation

1.1.1 Alternative bitcells

Technology scaling has contributed significantly to the improvement of density and performance of IC designs. However, continued scaling according to Moore's law faces several challenges in the sub-45 nm region, the main ones being increased leakage and process variations.

SRAM scaling is particularly challenging as it involves balancing trade-offs among yield, power, performance and area metrics. One example is the trade-off between functionality and power. Figure 1 shows the Read and Write Static Noise Margins using Predictive Technology Models (PTMs) [1]

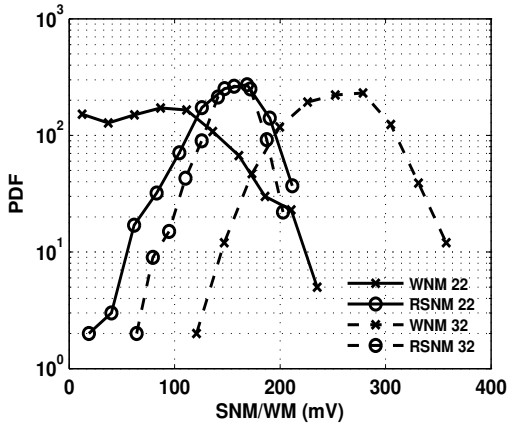

Figure 1: 22nm and 32nm noise margins using PTMs. The nominal voltage for 32 nm is 0.9 V and for 22 nm is 0.8 V.

We observe that as technology and supply voltage scale from 32 nm to 22 nm, not only are the mean noise margins reducing, but also the spread of the distributions is increasing. Since SRAM design takes into account the worst-case tails of these distributions, reduced functionality is clearly a barrier to scaling and to low voltage/low power operation. This is the functionality-power trade-off inherent in SRAM scaling.

Balancing such trade-offs becomes increasingly difficult in newer technologies, especially in light of worsened variation arising due to shrinking device dimensions in the bitcell. Thus, the best way to continue scaling SRAM beyond the 45nm node and to lower supply voltages for low-power operation is still an open question.

Problem 1: How do we better exercise the trade-offs involved in SRAM design to enable continued scaling?

To deal with this problem, we propose alternative single-ended bitcells that use asymmetric sizing as a knob to better exercise the trade-offs among power, performance, area and yield. Further, we propose a new sensing scheme that can be extended to any single-ended bitcell.

1.1.2 SRAM Design Automation

There are several solutions that combat SRAM design challenges, so that lower V_{MIN} can be achieved. For instance, at the circuit level, alternative bitcells such as the 8T [2] have been proposed. The 8T cell decouples the read and write operations which simplifies some of the trade-offs involved in SRAM design. Further, several read and write assist methods [3][4][5] have been proposed to improve noise margins. At the technology level, new materials and devices (e.g. high K-metal gate, Silicon on Insulator etc.) have been proposed to tackle the problems of leakage and variation that make SRAM design challenging. Architectural solutions such as ECC also help mask bit errors and bring down V_{MIN}.

Thus, there is a need for a methodology by which SRAM designers can easily explore the vast design space to decide which of these solutions or combination of solutions would result in a design that is optimal and meets required specifications in terms of the figures of merit of area, performance, power and yield.

Problem 2: How do we enable a rapid exploration of new and existing design techniques that combat SRAM scaling challenges?

To facilitate early and rapid SRAM design space exploration, we propose a technology-agnostic Virtual Prototyping tool (ViPro). It first provides a starting point for SRAM design in any technology by generating an optimal memory using characterized components from a built-in library. The designer can then explore the design space by changing the sub-components of the SRAM and allowing the tool to re-optimize the design. For instance, we can replace the conventional 6T bitcell used in the base-case SRAM with one of the alternative bitcells we propose. By allowing varying levels of detail and accuracy in specifying the sub-components, the tool allows a quick and early exploration of novel design techniques in terms of the high-level metrics and trade-offs among these metrics.

1.1.3 Dynamic Functional Margins

Finally, a big challenge in SRAM scaling is the increase in leakage power due to the scaling down of transistor threshold voltage (V_T). One way to combat this is to lower the supply voltage. However, this compromises the functional yield of the SRAM. An ability to predict how low the supply voltage of the SRAM can be pushed before failing to meet yield targets can help maximize power savings. Though it is possible to predict V_{MIN} based on static functionality and stability metrics [6], at high frequencies of operation such predictions may not be accurate due to the pessimistic or optimistic nature of static

metrics. Dynamic metrics are a better indicator of bitcell stability and functionality, but are harder to measure. Moreover, unlike static functionality metrics, there is no standard or widely accepted method of quantifying dynamic functionality or stability.

Problem 3: How are static measures of cell stability related to dynamic measures and will static measures be suitable to predict V_{MIN} for scaled SRAMs?

To deal with this problem, we propose to explore existing dynamic metrics and study their correlation with the static metrics.

1.2 Contributions

The high-level contributions of this dissertation will be –

- Alternative bitcell solutions that use asymmetric sizing as a knob to trade-off the critical metrics involved in SRAM design.
- A design methodology and CAD tool that enable early and iterative SRAM design space and trade-off exploration.
- Investigation of dynamic noise margin metrics and their correlation with static metrics.

2. Alternative Bitcells

2.1 Motivation

Increased variations reduce SRAM noise margins and oppose scaling of the conventional 6T SRAM bitcell (Figure 2) to lower V_{DD} and to new processes. Depending on the process and cell design, either Read Static Noise Margin (RSNM) or Write Noise Margin (WNM) tends to limit the V_{MIN}. In addition, embedded memories need to meet aggressive performance requirements. Since the 6T cannot meet stability and performance requirements simultaneously, alternative bitcells have been proposed. For example, the Nehalem processor [7] uses the 8T SRAM in the L1 and L2 caches to achieve high performance and adequate stability. In many designs, the area overhead of the 8T leads designers to explore other options for improving 6T stability such as read and write assists (e.g. [3][4][5]). The requirement for dense SRAM that has lower V_{MIN} while retaining stability and performance, especially in embedded memories for scaled technologies, suggests the need for an alternative bitcell to the 6T that provides a better tradeoff between area and these other critical metrics.

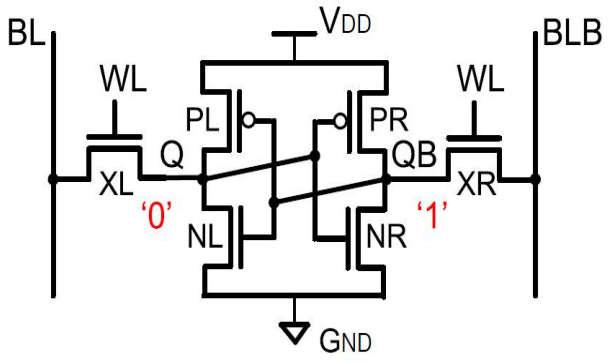

Figure 2: 6T bitcell

2.2 5T bitcell

We propose a 5T bitcell that closely mimics 6T access methods and that also uses a novel asymmetric sizing approach to increase RSNM and to provide an effective knob to trade-off performance, area, and other metrics.

Figure 3a shows the 5T bitcell schematic (a 6T bitcell missing one access transistor). Both read and write accesses occur identically to the 6T, except that they are single ended through the lone access device. Figure 3b shows the RSNM butterfly curve of a 5T bitcell obtained simply by dropping one access transistor from a conventional symmetrically sized 6T bitcell. One lobe of the curve is much smaller than the other due to the voltage-divider effect of N1 and N3. This lobe determines the RSNM of the "6T-like" 5T and is the same as the original 6T. Sizing the cross-coupled inverters in the bitcell asymmetrically skews the butterfly curve as indicated by the arrows in Figure 3b. Figure 3c shows the resulting increase in RSNM. This key insight provides the 5T bitcell with its beneficial features.

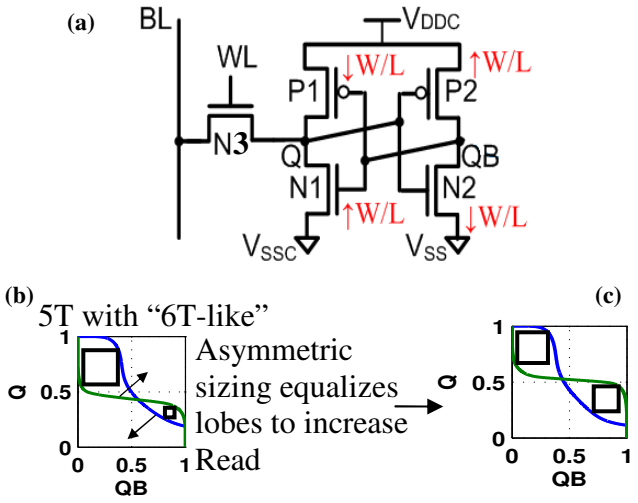

Figure 3: (a) 5T bitcell schematic. (b) Read Static Noise Margin (RSNM) for 5T with 6T-based sizing. (c) RSNM for 5T with asymmetric sizing

The main limitation of the 5T is degraded WNM compared with a 6T of iso-size, due to difficulty writing '1' through N3. We show that by collapsing V_{DDC} [8], we can solve this problem. As the timing waveforms in Figure 4 show, collapsing V_{DDC} weakens the cell feedback, enabling it to flip despite the weak '1' passed by N3.

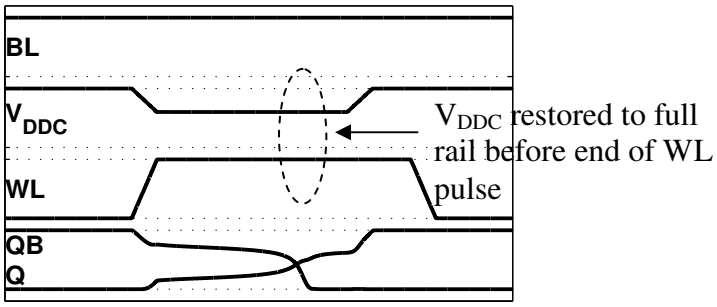

Figure 4: Timing for V_{DDC} collapse write-assist.

2.3 Asymmetric 6T bitcell

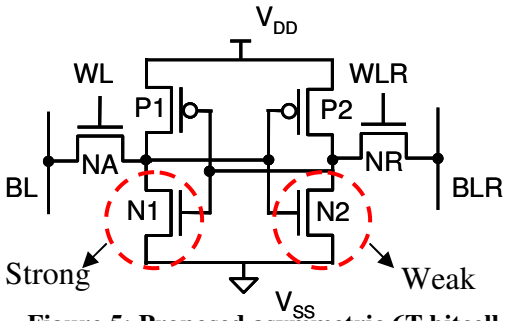

Figure 5: Proposed asymmetric 6T bitcell

We extend the asymmetric sizing concept to a 6T and propose an asymmetric single-ended 6T SRAM that improves both RSNM and WNM for the same bitcell area as a conventional 6T. An example implementation of the asymmetric 6T is depicted in Figure 5.

The improvement in noise margins is achieved using a single V_{DD}, without employing assist techniques that require multiple voltages, as was the case with the 5T bitcell described previously. This significantly improves the low-voltage robustness and consequently, V_{MIN}, without compromising on the writability of the cell.

Single-ended write is accomplished in two phases using dual word-lines. Figure 6a shows a generalized version of our proposed asymmetric 6T bitcell. The gate of the reset transistor, NR, is controlled by RSTG and the source by RSTS. RSTS is pre-charged high and RSTG is held low. As shown in Figure 6b, before a write, RSTS pulses low and RSTG pulses high and thus resets the cell to a '1'. RSTG and RSTS can be shared row-wise and column-wise, respectively, or vice-versa, and activated only for those rows and columns in the array that contain the accessed cell. This eliminates the need to reset an entire row and thus allows column muxing.

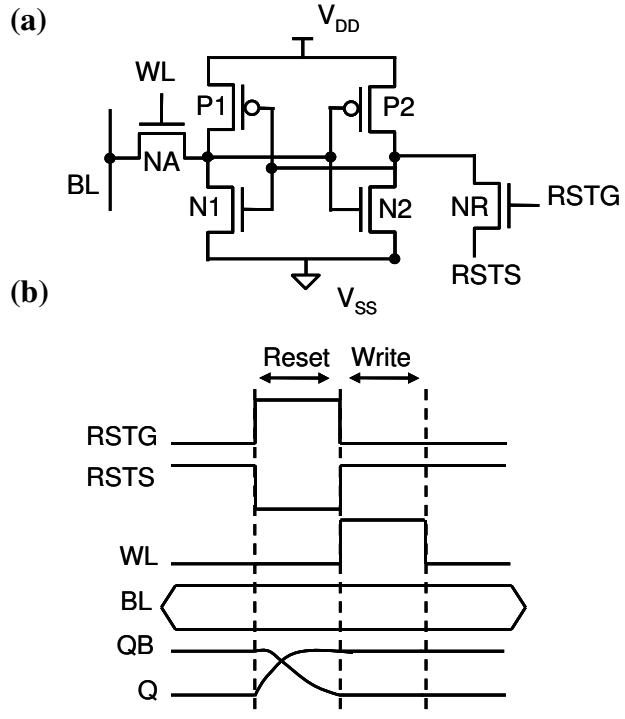

Figure 6: (a) Proposed cell reset scheme. (b) Timing diagram for cell reset and write

2.4 Single-ended Sensing

The improvement in I_{READ} due to asymmetric sizing for the proposed bitcells cannot be translated into an improvement in performance in terms of overall read access time. This is because logic-gate based full swing sensing that is conventionally used for single-

ended cells is slower than differential sensing [9] . Thus, we propose a sensing scheme for single-ended cells that can perform comparably to differential sensing.

Figure 7 depicts this scheme. The bitcell array is split in two halves and a conventional sense amplifier (SA) is placed in the middle. The top and bottom half of the array each contain a single row of weaker cells (with lower read current than the regular cell). This is used as the reference for the differential SA. A similar idea was described in [10] for single-ended sensing in a DRAM, where a storage capacitor with half the capacitance as the normal cell was used as the reference.

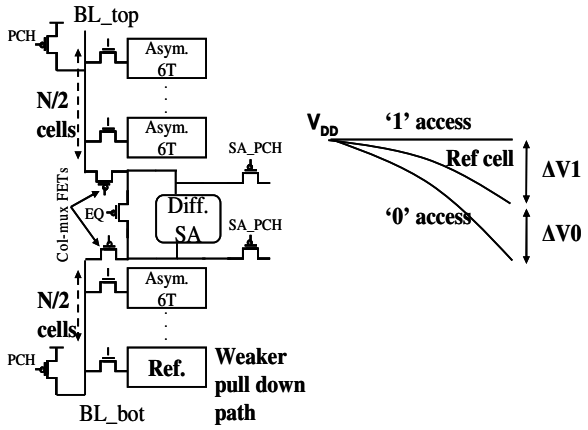

Figure 7: Single-ended sensing using a split BL.

The sensing works as follows. Without loss of generality, we assume that the cell being read is in the top half of the array. The wordline (WL) of the accessed row is activated, along with the WL of the reference row on the opposite side of the SA. If the cell being read stores a zero, it discharges the corresponding BL faster than the reference cell, as shown by the curve marked '0' access in Figure 7. On the other hand, if the cell being read stores a high value, the BL does not discharge, as indicated by the curve marked '1' access in the figure. The reference cell on the other side discharges as before. Thus, an appropriate differential is developed, which is quickly resolved by the SA. Note that the output of the SA would need to be inverted if the accessed cell was in the other half of the array.

In this scheme, in addition to an increase in read current, the bitline capacitance is reduced due to half the number of cells on the bitline. This helps compensate for the fact that the BL needs to be discharged twice as much as in the case of the conventional 6T (assuming the reference cell is half as strong as the normal cell) to generate the same net differential for the SA.

There are several other methods for single-ended sensing. One method is to use full-swing sensing with an inverter heavily skewed towards the PMOS. Full-swing sensing has been conventionally used for single-ended cells such as the 8T [2]. However, it is slower than small signal sensing using a SA, unless the bitlines are short and a hierarchical sensing scheme is used. Promising small-signal single-ended sensing

schemes such as the non-strobed regenerative SA (NSR-SA) have also been proposed [11].

2.5 Proposed Research

We have proposed the 5T bitcell earlier in [12]. That work presented a specific implementation of the asymmetric sizing approach that strengthened the pull down to achieve increased read current in addition to improved RSNM. In this dissertation, we will extend that work by generalizing the asymmetric sizing approach as a knob that allows us to trade-off the extra area not only for improved read current and RSNM, but also for lower leakage and lesser degradation of the WNM. We also provide new measurement results from a 5T SRAM fabricated in a commercial 45nm bulk CMOS technology.

Since the 6T cell has long been the workhorse of commercial memories, we will first compare the proposed asymmetric bitcells to the 6T in terms of power, performance, functionality/yield, and area. Further, since the 8T bitcell is now beginning to be used as an alternative to the conventional 6T, we will also compare the proposed cells with the 8T in terms of the same metrics. Simulation based comparisons will be substantiated with silicon measurements from a 45 nm 5T SRAM and a sub-45 nm asymmetric 6T SRAM. Finally, we will compare two existing single-ended sensing schemes, full-swing and NSR-SA, with our proposed scheme in terms of performance, power and array efficiency.

2.6 Contributions

- Present a 5T bitcell with a novel asymmetric sizing approach to increase RSNM over an iso-area 6T.
- Present an asymmetric 6T bitcell and a sizing approach that results in a higher mean and lower variability in the RSNM, WNM, and I_{READ} compared to an iso-area conventional symmetric 6T, allowing for improved V_{MIN}.
- Demonstrate 5T and asymmetric 6T SRAM in commercial CMOS technology and analyze the pros and cons of the alternative bitcells as compared to a conventional 6T and 8T.
- Demonstrate a pseudo-differential sensing scheme for this bitcell that can be extended to any single-ended bitcell, enabling improved performance while maintaining robustness.
- Compare single-ended sensing schemes with differential sensing, particularly for the proposed asymmetric bitcells.

2.7 Related Publications

1. **S. Nalam** and B. H. Calhoun**, "**Asymmetric Sizing in a 45nm 5T SRAM to Improve Read Stability over 6T", CICC, September 2009
2. **S. Nalam**, V. Chandra, C. Pietrzyk, R. Aitken, and B.H. Calhoun, " Asymmetric 6T SRAM with Two-phase Write and Split Bitline Differential Sensing for Low Voltage Operation", accepted at ISQED, 2010

3. Virtual Prototyping Tool (ViPro)

3.1 Motivation

While process scaling has enabled ever-larger embedded memories, scaling trends such as process variability, device leakage, soft error susceptibility, and interconnect delay make memory design increasingly difficult. In the face of such scaling effects, the best way to design SRAMs that are optimal in terms of global figures of merit (FoMs), which we define as energy, performance, area, and yield, at the 32nm process technology node and below, largely remains an open question. Researchers have proposed a number of techniques at the technology and circuit level to deal with problems such as variation and leakage, but they tend to address only certain individual components of the memory. A change in any one of the key memory circuits will alter the optimal circuit topologies, array partitioning, and architecture for the entire memory. This makes it difficult to evaluate a particular circuit technique without assessing the global benefits and overheads.

Though a back-of-the-envelope estimation of overheads and impact on SRAM global FoMs early in the design is possible, it largely depends on the assumptions made about the SRAM architecture and component circuits, some of which may not yet be designed. These assumptions can vary from designer to designer and lead to vastly different estimates about the impact of a particular circuit technique on SRAM global FoMs. Alternatively, the designer could create complete SRAM prototypes to evaluate each new technique, which impractically increases design time and reduces productivity. This makes it difficult for designers to make an accurate comparison of available design options and to choose one that results in an optimal design, especially in deep nanoscale processes. Thus, there is a need for a methodology through which designers can generate and evaluate prototypes at every step of the SRAM design process that account for process and circuit level issues in terms of global FoMs.

To address this problem, we propose a rapid Virtual Prototyping tool (ViPro), which enables early design space exploration by creating virtual prototypes of a complete SRAM macro, even when many design details are missing (hence “virtual”). As the design process proceeds, the prototypes become more accurate and complete. Thus, ViPro helps the designer do what he would want to do anyway (e.g. design space exploration), but much more efficiently, making it *design automation* in the truest sense.

There are a few memory design and FoM characterization tools available, but they do not support integrated process-circuit-system co-design like ViPro. At one end of the spectrum are architecture-level modeling tools like CACTI [13], used by computer architects to obtain quick estimates of SRAM access time, power, and area. CACTI 6.0 [14] facilitates high level design space exploration by using an optimization cost function that accounts for a user-weighted combination of delay, leakage, dynamic power, cycle time and area. ViPro also supports architectural exploration, but it differs from CACTI in two key ways. First, CACTI makes fixed assumptions regarding the circuits comprising the SRAM, so it optimizes at the architecture level only. ViPro allows designers to

generate circuit information (via simulation) specific to any given technology or to add/alter the underlying circuits. Thus, ViPro supports circuit-architecture co-design, which leads to better overall designs. Second, CACTI supports a limited set of process technologies and assumes ITRS [15] parameters for all its calculations. These assumptions may not be accurate, especially for advanced process nodes. ViPro uses a technology-agnostic simulation environment to characterize its circuit components in any process using SPICE before generating the virtual prototypes, so it uses accurate technology-specific circuit parameters for any process.

At the other end of the spectrum are transistor-level optimizers (e.g., [16][17]) that are good at choosing optimal device characteristics (e.g. W/L, V_T, V_{DD} etc.) for a given circuit topology, but are not helpful in choosing an optimal circuit topology or micro-architecture. In addition, since they are designed to thoroughly explore the design space under device and environmental variations, they are not suitable for quick, early design space exploration.

Finally, memory compilers (e.g. [18][19]) help generate memories based on user-defined parameters, but do not provide trade-off information or facilitate design space exploration and optimization. They are more a deliverable *from* memory design teams rather than a tool *for* memory design teams. ViPro fills the gap between these tools by providing an optimization and design space exploration tool for SRAM that supports circuit-system co-design.

3.2 Overview of ViPro

ViPro has four key features that make it a valuable tool for SRAM designers. First, ViPro generates a base-case SRAM prototype, one that falls in the middle of the trade-off space. Second, since the model allows the components to be described using varying levels of detail, designers can define and work on a complete prototype quite early in the design cycle. Third, ViPro can quickly re-optimize the design if a circuit component or the process models change. Finally, ViPro performs its own process characterization, and thus can be used with any process with defined SPICE/Spectre device models. In other words, it is technology-agnostic.

Figure 8 shows the structure of ViPro, which comprises two main blocks: a characterization engine (CE), and a hierarchical meta-compiler (HMC, ‘meta’ to distinguish it from a true compiler that produces complete final designs). The HMC implements an editable and flexible hierarchical model of the memory that allows a designer to define components of the memory with varying levels of detail and accuracy. The CE provides a technology-agnostic framework for generating data from simulation of those components, so that ViPro can operate in any process technology. The CE can also be used stand-alone as a Technology-agnostic Simulation Environment tool (TASE), not just to characterize SRAM components, but for circuit simulation in general.

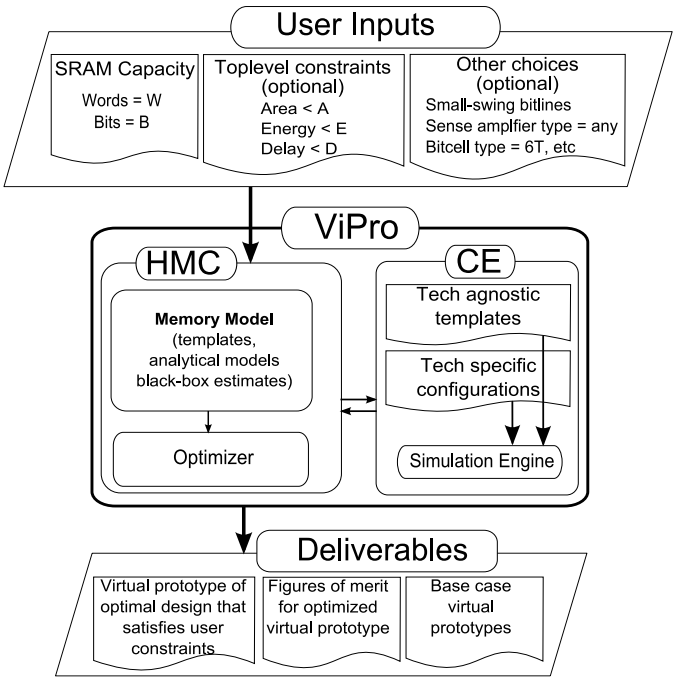

Figure 8: Top-level structure of Virtual Prototyper (ViPro). The characterization engine or Technology-agnostic simulation environment (CE/TASE) and the hierarchical meta-compiler (HMC) implement the two main features of ViPro - technology-agnosticism and a flexible hierarchical memory model.

3.2.1 Technology Agnostic Simulation Environment (TASE)

In [20], we propose a technology-agnostic simulation environment (TASE) that abstracts out the process dependencies from the simulation set-up using simulation templates. TASE then combines these templates with process-related data to produce the simulation-ready netlist and simulates it using Spectre. We incorporate TASE into ViPro for two purposes. First, we use it to characterize a technology or process through device-level simulations (e.g. I-V curves, FO4 delay, etc.). Second, TASE includes a user-expandable library of templates of SRAM components that can be characterized in terms of global (e.g. energy and delay) and component-specific FoMs (e.g. noise margins for a bitcell and offset for a sense amplifier). These templates also use Monte Carlo analysis to capture statistical data required for yield estimation. As new circuits are added to the template library over time, the CE incorporates a larger range of options that the HMC can then use for optimization. The technology-agnostic nature of TASE allows the CE to work with any technology that has a SPICE model file, which in turn makes ViPro usable for any technology.

3.2.2 Methodology

The designer provides the following inputs in a configuration file similar to CACTI.

- Process technology
- Top-level memory specifications – Capacity, word-size, supply voltage, operating temperature, etc.

- Constraints on metrics like energy, delay and area (optional – may be local or global)
- Component specifications (optional) – black-box estimates, analytical models or CE simulation templates

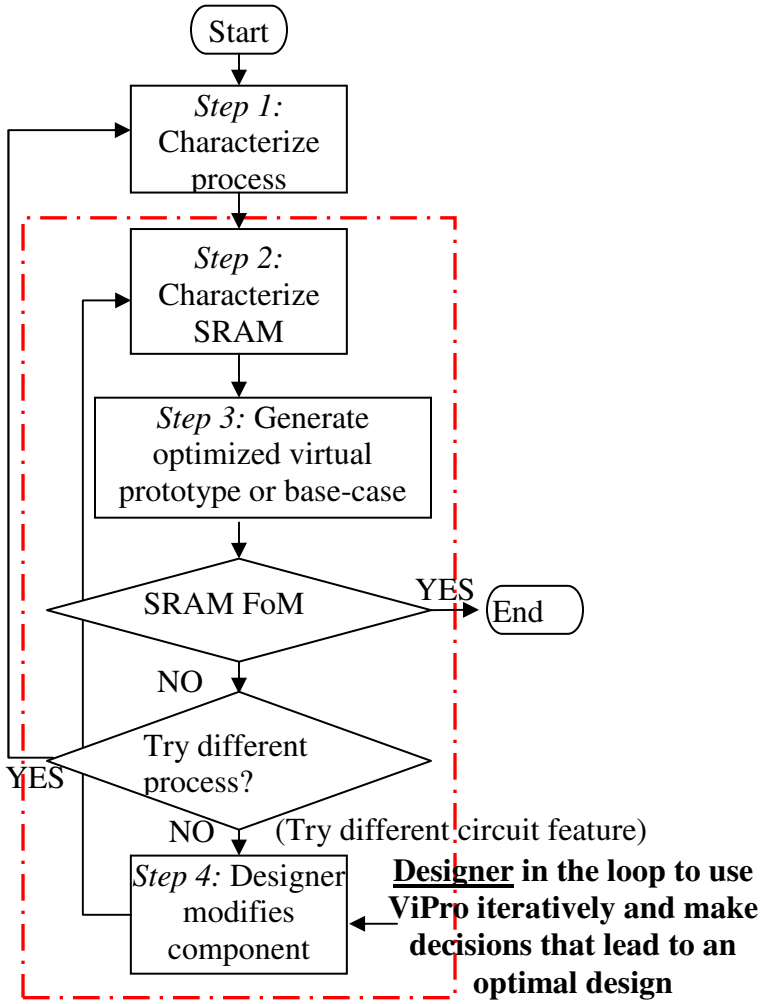

Figure 9: Methodology of using virtual prototypes for SRAM design

Figure 9 shows the steps involved in using ViPro for generating virtual prototypes. The first step (once per technology) is to characterize the process technology through device-level simulations. In the second step, the CE characterizes components from the library (once per technology). Any components unavailable from the CE must be defined using black-box estimates, analytical models, or new templates added to the CE library. As more components are added to the library, the accuracy and scope of the virtual prototypes improves. In the third step, using existing components and built-in analytical models, ViPro can generate an optimized base-case prototype that provides a convenient starting point for a designer interested in creating a more optimized custom design. He can explore different circuit options (e.g. assist techniques, alternative bitcells, etc.) to further optimize his/her design. By changing the specification of one or more components and running ViPro iteratively, the designer can steer the effort towards an optimal design.

Alternatively, the designer can exploit the technology-agnostic nature of ViPro to compare prototypes for different process or device options. This kind of comparison is especially useful for many "fab-less" companies that have to choose between different processes for their design. Thus, for example, when porting an existing design to a new technology, the designer can quickly see how the optimum configuration of the design changes. This technology agnostic feature also allows for process-circuit co-design, since the optimal circuit and architecture selections will change in response to process alterations.

A key insight here is that the designer is an integral part of the tool flow. As the designer runs the tool multiple times and compares several virtual prototypes, his/her understanding of the trade-offs involved in the design increases. Thus, he finalizes more and more components, which improves the accuracy of the prototype. Ultimately, as the design nears completion, all the components in the memory are specified in terms of circuit netlists from the template library of the CE, and ViPro becomes closer to an actual SRAM compiler.

3.3 Proposed Research

First, we propose to fabricate a test chip in commercial 65 nm bulk CMOS technology to validate the accuracy of the virtual prototypes generated by ViPro. The test chip will have several SRAM blocks with different design parameters (e.g. aspect ratio). The blocks will be semi-automatically generated by ViPro using a technology-agnostic SKILL based flow that will require manual development of individual SRAM sub-blocks. Delay and power measurements from these blocks will indicate whether the fabricated SRAM blocks lie on the same Pareto Energy-Delay curve as predicted by ViPro.

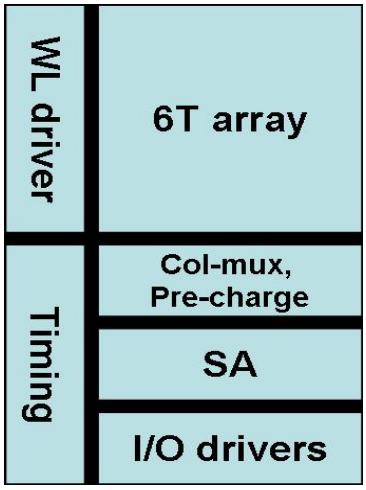

Figure 10: SRAM architecture assumed by ViPro

For this experiment, to simplify the problem, we will fix most of the design parameters and have only a limited set of variables in the optimization problem. Figure 10 shows the assumed architecture of the blocks. Only the WL and the column drivers will be optimized by ViPro and the other components like the timing and the SA blocks will

have a fixed design. The top-level knob for optimization will be the aspect ratio of the bitcell array. The optimal solution will be determined by a brute-force search through all possible (limited) combinations.

Next, we will shift our focus to increasing the utility of the tool. For this, we will first add more sub-blocks and add multiple types of each sub-block (e.g. different types of SAs) to TASE. Then, we will increase the number of local and global variables/knobs in the optimization problem and plug in a commercial optimization tool (e.g. MOSEK [21]) to handle the increased complexity. We will also compare the FoMs of the prototypes generated by ViPro with those estimated by CACTI.

Finally, with ViPro in a usable form, we will demonstrate two examples of how it can be used for early design space exploration. In the first example, we will look at how different bitcells affect the optimal SRAM design in different technologies. For this purpose, we propose to experiment with the conventional 6T, 8T, 5T, and the asymmetric 6T. In the second example, we will look at how different read and write assist techniques affect the optimal SRAM design in different technologies.

3.4 Contributions

- ViPro
    - Quickly generates virtual prototypes and top-level SRAM FoM trade-off data at every step of the design process.
    - Provides a starting point design in any technology.
    - Re-optimizes the design to reflect changes at the process, circuit, and architectural levels, while not requiring completed description of the component circuits and involving the designer in an iterative process.

- TASE:
    - Automates *groups* of related simulations across technologies
    - Preserves and exposes designer intent and underlying tradeoffs, allowing rapid redesign and adjustment to unforeseen issues.
    - Increases productivity of an iterative design process.
    - Expands and “learns” with use as it captures increased levels of the designer knowledge and intent.

3.5 Related Publications

1. M. Bhargava, **S. Nalam**, B. H. Calhoun, K. Mai, "An SRAM Prototyping Tool for Rapid Sub-32nm Design Exploration and Optimization", TECHCON, September 2009
2. **S. Nalam**, M. Bhargava, K. Ringgenberg, K. Mai, and B. H. Calhoun, "A Technology-Agnostic Simulation Environment (TASE) for Iterative Custom IC Design across Processes", ICCD, October 2009
3. **S. Nalam,** M. Bhargava, K. Mai, and B.H. Calhoun, “Virtual Prototyper (ViPro): An early design space exploration and optimization tool for SRAM designers”, accepted at DAC 2010

4. Dynamic Noise Margin Analysis

4.1 Motivation

Increasing power consumption is a major problem in modern electronics. Low-voltage operation provides power-savings or longer battery life depending on the application. However, reducing the voltage decreases noise margins, particularly during read and write. In addition, the use of minimum-sized devices in the 6T bitcell exacerbates the variability of the already low noise margins. This impacts the functional yield and restricts the extent to which the supply voltage can be lowered for achieving low-power operation (V_{MIN}). Further, the ever-increasing capacity of on-chip caches and other SRAM based memories makes achieving yield targets even more challenging. Thus, an accurate and quick prediction of V_{MIN} and the yield achievable at a particular supply voltage and temperature is needed to accelerate the exploration of the trade-offs among performance, power, and yield, and contribute to designing an optimum SRAM for a given application.

Existing attempts at prediction of V_{MIN} for active operation and yield estimation are based on static metrics for noise margin during read, write [5]. However, static metrics tend to be optimistic for write and pessimistic for read, since the cell disturbance is considered to be of infinite duration. Thus the problem with using V_{MIN} predicted by static methods is that could lead to reduced power savings or failure to meet yield targets. We propose to correct this by basing V_{MIN} prediction and yield estimation on dynamic noise margin metrics that take into account the transient behavior of the SRAM bitcell.

4.2 Existing Approaches

SRAM yield may be divided into two categories, (1) hard fails, those attributable to defects, and (2) soft fails. Soft fails are voltage, temperature and timing dependent fails resulting from one of the following four modes: (1) failure to write, (2) failure to read (insufficient signal developed on the BL to set the sense amp), (3) stability upset and (4) data retention. These four failure modes are associated with a distribution tail stemming from variation sources.

For hold/read stability failure, butterfly based static noise margin (SNM) is widely used [22]. Recently, N-curve based static current margin (I_{CRIT}) is also proposed to measure cell read stability [23]. In addition to SNM and I_{CRIT}, static voltage margins on BL [24] or WL [25] gain popularity for quantifying cell write ability. Compared with dynamic margin, static margin is easy to measure. However, it is always pessimistic for read because it measures maximum tolerable dc noise, which is often smaller than dynamic margin (DM). By a similar analogy, static metrics are always optimistic for write. Since the real read/write operation is performed in a dynamic fashion, i.e. with the timing constraint, DM is the true indicator of functional margin. This fact has lead to several recent publications exploring dynamic margins for SRAM [26][27][28][29][30]. We have proposed a dynamic write margin (T_{CRIT}) that defines the minimum WL pulse width for the cell trajectory to cross over the cell's separatrix line [31]. This separatrix-line based

dynamic margin is also used for read stability. Nho, et al, [29] have proposed a Monte-Carlo simulation method to estimate read access yield combining the statistics of SRAM cells and surrounding circuits such as SA.

4.3 Dynamic Write Margin

4.3.1 T_{CRIT} Writability Metric

In order to evaluate the write ability of a cell more precisely, a metric which takes into account the dynamic write behavior must be used. On a successful write, the storage node and its complement (Q and QB) cross over and eventually settle at V_{DD} and 0V, respectively. The WL pulse width (T_{WL}) determines whether or not the two waveforms cross and the write is successful. We propose to use the minimum WL pulse width (T_{CRIT}) for the cell to flip ultimately to the correct new state as a metric for dynamic write margin.

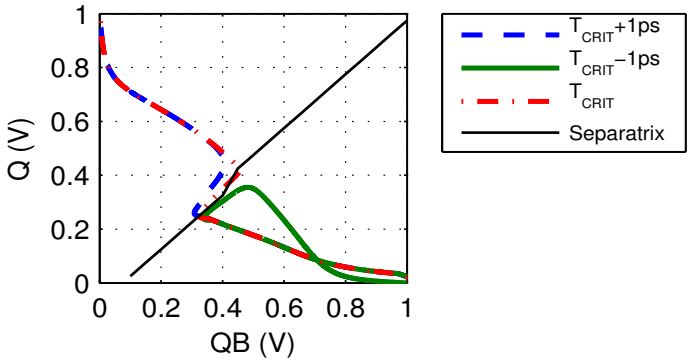

Figure 11: Trajectories for writing '1' (starting point (QB, Q) = (1, 0)) when T_{WL} is equal to T_{CRIT} and T_{CRIT}±1ps.

Figure 11 shows the trajectories of Q and QB for writing '1' when T_{WL} is equal to T_{CRIT}−1ps, T_{CRIT}, and T_{CRIT}+1ps. The thinner solid line is the separatrix for this cell, which is the cell with the largest T_{CRIT} out of 1000 M-C simulations. The trajectories all overlap each other as they approach the separatrix, but they diverge at that point because the WL pulses end at slightly different times. The two trajectories for $T_{WL} \geq T_{CRIT}$ then overlap again as they converge to the newly written value, but the trajectory for $T_{WL} < T_{CRIT}$ falls back to the starting state. This simulation clearly shows that T_{CRIT} is the WL pulse width that causes the state of the cell (QB, Q) to cross over the separatrix when the WL drops to 50% of V_{DD}. Notice that variation has pushed the separatrix off of the line Q=QB. As in this cell, when we write '0' instead (trajectory starts from (QB, Q) = (0, 1)), the voltages at Q and QB can actually cross, but, if the trajectory does not cross the separatrix, the write can still eventually fail.

4.3.2 Correlation of T_{CRIT} with Static Writability Metrics

Static measures of write margin are convenient and fast for simulation and testing. If they accurately reflect the dynamic behavior of the write operation, we can use them to

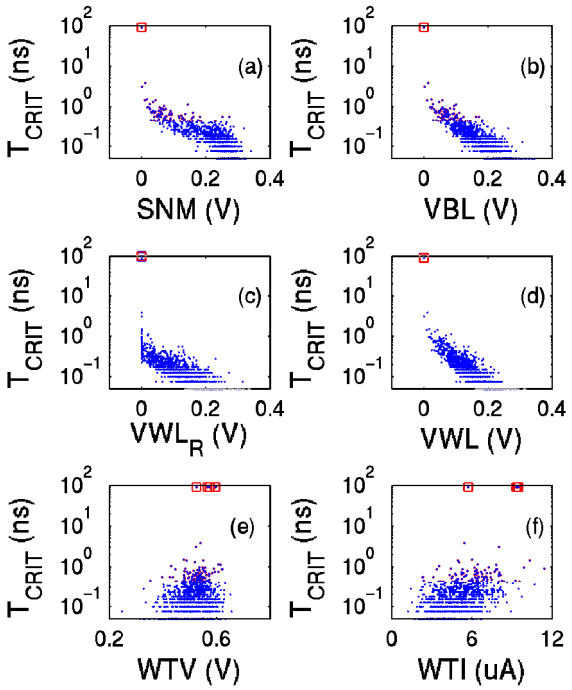

Figure 12: Correlation between each static metric and T_{CRIT} at V_{DD}=0.6V. Points highlighted with square actually never flip ($T_{CRIT}=\infty$); their values are assumed to 90ns for display.

identify cells that will limit write ability in the presence of variation and at lower operating voltages.

Figure 12 shows the correlation between T_{CRIT} and each static metric at V_{DD}=0.6V. The correlation plots reveal that the N-curve metrics [32], WTV and WTI, have a poor correlation with T_{CRIT}. A higher WTI or WTV value is supposed to imply a weaker write ability, which should indicate a higher T_{CRIT}. However, in Figure 12e-f, those points with highest T_{CRIT} correspond to a wide range of WTI and WTV values. WTI and WTV are measured in the context of a read operation, so they should be considered as metrics for read stability instead of write ability. Write ability has a strong dependence on $V_{T\text{-}X1}$, the threshold voltage of the access transistor for node '1', because the corresponding BL is 0 during write and a large amount of current would flow through it to discharge the node '1'.

Table 1: Correlation coefficient (CC) between each write margin and $\Delta V_{T\text{-}X1}$ at V_{DD}=0.6V

	T_{CRIT}	SNM	VBL	VWL_R	VWL	WTV	WTI
CC	0.60	-0.54	-0.61	-0.60	-0.65	-0.15	-0.03

Table 1 shows the correlation coefficient between static/dynamic write margin metrics and $V_{T\text{-}X1}$ variation ($\Delta V_{T\text{-}X1}$). Most of them have a correlation coefficient with large magnitude (e.g. 0.60 for T_{CRIT}), which verifies that $V_{T\text{-}X1}$ has a strong impact on write ability. WTI and WTV have the least correlation with $V_{T\text{-}X1}$ because both bitlines are

clamped at V_{DD} in the simulation for these metrics. Therefore, WTI and WTV are not good metrics for write ability. The VWL, VWL_R, VBL, and SNM metrics all show a good match for the never-flipped case (those points highlighted with square), and a good correlation to T_{CRIT}, especially at the worst case tail.

4.4 Proposed Research

Unlike static metrics, there is no widely accepted dynamic metric for cell readability or writability. So, we will first survey existing techniques for measuring dynamic noise margin for read and write. After studying the correlation between the dynamic and static metrics (as we have shown for T_{CRIT}), we will pick one that correlates the best with the conventional static metrics and is also relatively easy to measure. We will then aim to identify the sources of non-correlation between the chosen dynamic and static metrics by starting with the simplest possible transistor models with minimal parameters.

The proposed research could expose potential issues with using static metrics as indicators of cell readability/writability, especially for current nano-scale technologies. For instance, the number of bits on the BL, their parasitic leakage paths, and the ability of the design to enable arrival of the data before WL selects during a write, contribute toward the dynamic writability of the cell. These factors are not captured by static metrics which essentially perform dc analysis on the bitcell alone. Thus V_{MIN} predictions based on static metrics could be inaccurate. Understanding the relation between static metrics and the dynamic readability/writability of the cell can help in either coming up with a dynamic-metric-based method of predicting V_{MIN} or an improvement in the accuracy of V_{MIN} prediction based on static metrics.

4.5 Contributions

- A comprehensive study of existing dynamic stability metrics.
- Identification of best dynamic metric in terms of measurability and correlation with static metrics.
- Study of correlation between chosen dynamic and static metrics.
- Identification of causes of non-correlation between chosen dynamic and static metrics.

4.6 Related Publications

1. J. Wang, **S. Nalam**, and B. H. Calhoun, "Analyzing Static and Dynamic Write Margin for Nanometer SRAMs", ISLPED, August 2008

5. Timeline of Research Tasks

Table 2 below lists the tasks, status, and relevant publications for each research goal.

Table 2: Research tasks and status

Subject	#	Task Description	Status/Target	Pubs
Alternative Bitcells				
5T	1	Compare with 6T in terms of power, performance, stability/functionality, and area	Done	[SVN3] [SVN8]
	2	Compare with 8T in terms of above metrics	Sept 2010	[SVN11]
	3	Measurements from 45nm chip	Done	[SVN3] [SVN8]
Asymmetric 6T	4	Compare with 6T	Done	[SVN5]
	5	Compare with 8T	Sept 2010	[SVN11]
	6	Measurements from sub-45nm chip	Aug 2010	[SVN12]
Single-ended sensing	7	Investigate single-ended sensing for 5T and asymmetric 6T	March 2010 (5T) Oct 2010 (Asym. 6T)	[SVN8] [SVN12]
Virtual Prototyping Tool	1	Technology-agnostic simulation environment (TASE)	Done	[SVN4]
	2	Hierarchical model in MATLAB	Done	[SVN2] [SVN7]
	3	Technology-agnostic SRAM generator	April 2010	[SVN13]
	4	Tape-out 65 nm chip with ViPro-generated SRAM		
	5	Add simulation templates for SRAM components to TASE	May 2010	
	6	Incorporate suitable optimization tool	Nov 2010	[SVN14]
	7	Compare with CACTI	Jan 2011	[SVN13]
	8	Measurements from 65 nm chip		
	9	Demonstrate design space exploration	April 2011	[SVN14]
Dynamic stability	1	Survey existing dynamic metrics	June 2010	[SVN15]
	2	Relation between static and dynamic metrics, investigate sources of non-correlation	Aug 2010	[SVN1] [SVN15]
Write-up	1	Dissertation writing	May 2011	

6. Publications

6.1 Current Publications

[SVN1] J. Wang, **S. Nalam**, and B. H. Calhoun, "Analyzing Static and Dynamic Write Margin for Nanometer SRAMs", ISLPED, August 2008

[SVN2] M. Bhargava, **S. Nalam**, B. H. Calhoun, K. Mai, "An SRAM Prototyping Tool for Rapid Sub-32nm Design Exploration and Optimization", TECHCON, September 2009

[SVN3] **S. Nalam** and B. H. Calhoun**, "**Asymmetric Sizing in a 45nm 5T SRAM to Improve Read Stability over 6T", CICC, September 2009

[SVN4] **S. Nalam**, M. Bhargava, K. Ringgenberg, K. Mai, and B. H. Calhoun, "A Technology-Agnostic Simulation Environment (TASE) for Iterative Custom IC Design across Processes", ICCD, October 2009

[SVN5] **S. Nalam**, V. Chandra, C. Pietrzyk, R. Aitken, and B.H. Calhoun, " Asymmetric 6T SRAM with Two-phase Write and Split Bitline Differential Sensing for Low Voltage Operation", accepted at ISQED, 2010

[SVN6] R. Mann, **S. Nalam**, J. Wang, and B. H. Calhoun, "Limits of Bias Based Assist Methods in Nano-Scale 6T SRAM", accepted at ISQED, 2010

[SVN7] **S. Nalam,** M. Bhargava, K. Mai, and B.H. Calhoun, "Virtual Prototyper (ViPro): An early design space exploration and optimization tool for SRAM designers", accepted at DAC 2010

6.2 Anticipated Publications

[SVN8] **S. Nalam** and B. H. Calhoun**, "**Asymmetric Sizing in a 5T SRAM for Improved Read Stability", to be submitted to JSSC

[SVN9] J. Wang, **S. Nalam**, Z. Qi, R. W. Mann, M. Stan, and B. H. Calhoun, "SRAM Vmin/Yield Improvement Using Variation-Aware BTI Stress", submitted to VLSI Symposium

[SVN10] R. W. Mann, J. Wang, **S. Nalam**, S. Khanna, and B. H. Calhoun, "Impact of circuit assist methods on both margin and performance in 6T SRAM", submitted to Journal of Solid-State Electronics

[SVN11] Comparison of asymmetric bitcells with 8T bitcell

[SVN12] Measured results for asymmetric 6T

[SVN13] Validation of ViPro and comparison with CACTI

[SVN14] Design space exploration using ViPro

[SVN15] Comparison and analysis of dynamic functionality metrics

References:

[1] Wei Zhao; Yu Cao; , "New generation of predictive technology model for sub-45nm design exploration," *Quality Electronic Design, 2006. ISQED '06. 7th International Symposium on* , vol., no., pp.6 pp.-590, 27-29 March 2006

[2] Chang, L.; Montoye, R.K.; Nakamura, Y.; Batson, K.A.; Eickemeyer, R.J.; Dennard, R.H.; Haensch, W.; Jamsek, D.; , "An 8T-SRAM for Variability Tolerance and Low-Voltage Operation in High-Performance Caches," *Solid-State Circuits, IEEE Journal of* , vol.43, no.4, pp.956-963, April 2008

[3] Ohbayashi, S.; Yabuuchi, M.; Nii, K.; Tsukamoto, Y.; Imaoka, S.; Oda, Y.; Yoshihara, T.; Igarashi, M.; Takeuchi, M.; Kawashima, H.; Yamaguchi, Y.; Tsukamoto, K.; Inuishi, M.; Makino, H.; Ishibashi, K.; Shinohara, H.; , "A 65-nm SoC Embedded 6T-SRAM Designed for Manufacturability With Read and Write Operation Stabilizing Circuits," *Solid-State Circuits, IEEE Journal of* , vol.42, no.4, pp.820-829, April 2007

[4] Pilo, H.; Barwin, J.; Braceras, G.; Browning, C.; Burns, S.; Gabric, J.; Lamphier, S.; Miller, M.; Roberts, A.; Towler, F.; , "An SRAM Design in 65nm and 45nm Technology Nodes Featuring Read and Write-Assist Circuits to Expand Operating Voltage," *VLSI Circuits, 2006. Digest of Technical Papers. 2006 Symposium on* , vol., no., pp.15-16

[5] Bhavnagarwala, A.J.; Kosonocky, S.; Radens, C.; Yuen Chan; Stawiasz, K.; Srinivasan, U.; Kowalczyk, S.P.; Ziegler, M.M.; , "A Sub-600-mV, Fluctuation Tolerant 65-nm CMOS SRAM Array With Dynamic Cell Biasing," *Solid-State Circuits, IEEE Journal of* , vol.43, no.4, pp.946-955, April 2008

[6] Wang J. and Calhoun B. H., "Minimum Supply Voltage and Yield Estimation for Large SRAMs under Parametric Variations", submitted to TVLSI

[7] Kumar, R.; Hinton, G.; , "A family of 45nm IA processors," *Solid-State Circuits Conference - Digest of Technical Papers, 2009. ISSCC 2009. IEEE International* , vol., no., pp.58-59, 8-12 Feb. 2009

[8] Yamaoka, M.; Maeda, N.; Shinozaki, Y.; Shimazaki, Y.; Nii, K.; Shimada, S.; Yanagisawa, K.; Kawahara, T.; , "Low-power embedded SRAM modules with expanded margins for writing," *Solid-State Circuits Conference, 2005. Digest of Technical Papers. ISSCC. 2005 IEEE International* , vol., no., pp.480-611 Vol. 1, 10-10 Feb. 2005

[9] Yibin Ye; Khellah, M.; Somasekhar, D.; De, V.; , "Evaluation of differential vs. single-ended sensing and asymmetric cells in 90 nm logic technology for on-chip caches," *Circuits and Systems, 2006. ISCAS 2006. Proceedings. 2006 IEEE International Symposium on* , vol., no., pp.4 pp.-966

[10] Poulton, J.; , "An embedded DRAM for CMOS ASICs," *Advanced Research in VLSI, 1997. Proceedings., Seventeenth Conference on* , vol., no., pp.288-302, 15-16 Sep 1997

[11] Verma, N.; Chandrakasan, A.P.; , "A High-Density 45 nm SRAM Using Small-Signal Non-Strobed Regenerative Sensing," *Solid-State Circuits, IEEE Journal of* , vol.44, no.1, pp.163-173, Jan. 2009

[12] Nalam S., "A 5T SRAM with improved read stability and variation tolerance over 6T," M.S. thesis, Univ. of Virginia, 2008.

[13] Shivakumar P. and Jouppi N. P., "*Cacti 3.0: An Integrated Cache Timing, Power and Area Model*," Tech. Rep. Western Research Lab (WRL) Research Report, 2001/2

[14] Muralimanohart, N.; Balasubramonian, R.; Jouppi, N.; , "Optimizing NUCA Organizations and Wiring Alternatives for Large Caches with CACTI 6.0," *Microarchitecture, 2007. MICRO 2007. 40th Annual IEEE/ACM International Symposium on* , vol., no., pp.3-14, 1-5 Dec. 2007

[15] Semiconductor Industries Association. International Technology Roadmap for Semiconductors. http://www.itrs.net/, 2006 update.

[16] Xiaoliang Bai; Visweswariah, C.; Strenski, P.N.; Hathaway, D.J.; , "Uncertainty-aware circuit optimization," *Design Automation Conference, 2002. Proceedings. 39th* , vol., no., pp. 58- 63, 2002

[17] Sundararajan, V.; Sapatnekar, S.S.; Parhi, K.K.; , "Fast and exact transistor sizing based on iterative relaxation," *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on* , vol.21, no.5, pp.568-581, May 2002

[18] Chakraborty, K.; Kulkami, S.; Bhattacharya, M.; Mazumder, P.; Gupta, A.; , "A physical design tool for built-in self-repairable RAMs," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* , vol.9, no.2, pp.352-364, Apr 2001

[19] Chandna A., "GaAs MESFET static RAM design for embedded applications," Ph.D. dissertation, Univ. Michigan, 1995.

[20] Nalam S., et al., "A Technology-Agnostic Simulation Environment (TASE) for Iterative Custom IC Design across processes", *ICCD*, 2009.

[21] "The MOSEK Optimization Software", http://www.mosek.com/

[22] Seevinck, E.; List, F.J.; Lohstroh, J.; , "Static-noise margin analysis of MOS SRAM cells," *Solid-State Circuits, IEEE Journal of* , vol.22, no.5, pp. 748- 754, Oct 1987

[23] Wann, C.; Wong, R.; Frank, D.J.; Mann, R.; Shang-Bin Ko; Croce, P.; Lea, D.; Hoyniak, D.; Yoo-Mi Lee; Toomey, J.; Weybright, M.; Sudijono, J.; , "SRAM cell design for stability methodology," *VLSI Technology, 2005. (VLSI-TSA-Tech). 2005 IEEE VLSI-TSA International Symposium on* , vol., no., pp. 21- 22, 25-27 April 2005

[24] Kevin Zhang; Bhattacharya, U.; Zhanping Chen; Hamzaoglu, F.; Murray, D.; Vallepalli, N.; Yih Wang; Bo Zheng; Bohr, M.; , "A 3-GHz 70-mb SRAM in 65-nm CMOS technology with integrated column-based dynamic power supply," *Solid-State Circuits, IEEE Journal of* , vol.41, no.1, pp. 146- 151, Jan. 2006

[25] Gierczynski, N.; Borot, B.; Planes, N.; Brut, H.; , "A New Combined Methodology for Write-Margin Extraction of Advanced SRAM," *Microelectronic Test Structures, 2007. ICMTS '07. IEEE International Conference on* , vol., no., pp.97-100, 19-22 March 2007

[26] Huang, G.M.; Wei Dong; Yenpo Ho; Peng Li; , "Tracing SRAM separatrix for dynamic noise margin analysis under device mismatch," *Behavioral Modeling and Simulation Workshop, 2007. BMAS 2007. IEEE International* , vol., no., pp.6-10, 20-21 Sept. 2007

[27] Wei Dong; Peng Li; Huang, G.M.;, "SRAM dynamic stability: Theory, variability and analysis," *Computer-Aided Design, 2008. ICCAD 2008. IEEE/ACM International Conference on* , vol., no., pp.378-385, 10-13 Nov. 2008

[28] Bin Zhang; Arapostathis, A.; Nassif, S.; Orshansky, M.; , "Analytical Modeling of SRAM Dynamic Stability," *Computer-Aided Design, 2006. ICCAD '06. IEEE/ACM International Conference on* , vol., no., pp.315-322, 5-9 Nov. 2006

[29] Hyunwoo Nho; Sei-Seung Yoon; Wong, S.S.; Seong-Ook Jung; , "Numerical Estimation of Yield in Sub-100-nm SRAM Design Using Monte Carlo Simulation," *Circuits and Systems II: Express Briefs, IEEE Transactions on* , vol.55, no.9, pp.907-911, Sept. 2008

[30] Khalil, D.E.; Khellah, M.; Nam-Sung Kim; Ismail, Y.; Karnik, T.; De, V.K.; , "Accurate Estimation of SRAM Dynamic Stability," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on* , vol.16, no.12, pp.1639-1647, Dec. 2008

[31] Wang J., Nalam S., and Calhoun B. H., “Analyzing static and dynamic write margin for nanometer SRAMs,” *ISLPED*, pp. 129–134, Aug. 2008.

[32] Grossar, E.; Stucchi, M.; Maex, K.; Dehaene, W.; , "Read Stability and Write-Ability Analysis of SRAM Cells for Nanometer Technologies," *Solid-State Circuits, IEEE Journal of* , vol.41, no.11, pp.2577-2588, Nov. 2006